

EDGE Computing for Mobile Networks 2019

Abstract:

Edge Computing is coming to the mobile network, and this report investigates who will invest in Edge Computing infrastructure, and how the business model will evolve. The analysis includes estimates of revenue for mobile operators in key applications ranging from gaming and AR/VR to industrial IoT, robotics, mining, drones, and automotive applications. The revenue is segmented by latency level to illustrate where the best investment strategy will be for Cloud providers and mobile telcos. Forecasts for data centers and edge servers are included for the global market through 2024.



MOBILE EXPERTS

J. Madden

April 2019

Edge Computing for Mobile Networks 2019

1	EXECUTIVE SUMMARY	6
2	MARKET DRIVERS AND CHALLENGES	9
	THE APPS	10
	LOWER LATENCY	10
	DISTRIBUTED CAPACITY	10
	THE EDGE COMPUTING BUSINESS MODEL.....	12
	...MOBILE OPERATORS AND A “TELCO CLOUD”	12
	...OPERATOR-HOSTED CLOUD	13
	...REIT HOSTED DATA CENTERS	13
	...WEB-SCALE PLAYERS DEPLOYING THEIR OWN DATA CENTERS	14
	THE ECONOMICS OF EDGE COMPUTING: DOES IT SAVE MONEY?.....	14
	THE ECONOMICS OF EDGE COMPUTING: DOES IT ADD REVENUE?	17
3	REVENUE OUTLOOK	19
	MOBILE OPERATOR REVENUE.....	19
4	DATA CENTER AND SERVER OUTLOOK	21
	DEFINITIONS.....	21
	DATA CENTER FORECAST	23
	EDGE SERVER FORECAST.....	26
5	KEY COMPANIES.....	30
	CLOUD PROVIDERS	30
	ALIBABA	30
	AWS (AMAZON WEB SERVICES).....	30
	GOOGLE (GCP, GOOGLE CLOUD PLATFORM)	30
	IBM	30
	LIMELIGHT NETWORKS	30
	MICROSOFT AZURE	30
	PACKET	30
	DATA CENTER/COLOCATION	30
	AMERICAN TOWER.....	30
	CBRE	30
	COLT	30
	COMPASS	30
	CROWN CASTLE	30
	CYRUS ONE.....	30
	DIGITAL FORTRESS.....	30
	DIGITAL REALTY	30
	EDGECONNEX	30
	EQUINIX	30
	GLOBAL SWITCH	30
	HITACHI	30
	ISEEK.....	30
	KVH.....	30
	KDDI.....	30
	KEPPEL.....	30

NEXTDC.....	30
RACKSPACE	31
SRP.....	31
TATA	31
ENCLOSURES/MICRO DATA CENTERS.....	31
BASELAYER.....	31
DARTPOINTS	31
DATABANK	31
EDGEMICRO	31
MODULAR LIFE SOLUTIONS/EDGEPOD	31
STULZ	31
VAPOR IO.....	31
VERTICAL BRIDGE	31
CDNS	31
AKAMAI	31
CDN77	31
CLOUDFLARE	31
FASTLY	31
GOOGLE CLOUD CDN	31
HIGHWINDS	31
IMPERVA ENCAPSULA.....	31
QWILT	31
STACKPATH.....	31
HARDWARE OEMS.....	31
ADLINK.....	31
CISCO	31
DELL	31
ERICSSON	31
HPE	32
HUAWEI	32
INTEL	32
NOKIA.....	32
NVIDIA	32
SAMSUNG.....	32
ZTE	32
SOFTWARE SUPPLIERS.....	32
AFFIRMED NETWORKS	32
ALTIOSTAR	32
ARICENT	32
CLEARBLADE.....	32
CPLANE.AI	32
JUNIPER NETWORKS	32
MAVENIR	32
QUORTUS.....	32
RIGADO.....	32
RTI (REAL-TIME INNOVATIONS).....	32
SAGUNA	32
VASONA NETWORKS.....	32
VMWARE	32
6 METHODOLOGY.....	33

CHARTS

Chart 1: Forecasted Mobile Operator EC-enabled Revenue thru 2025.....	7
Chart 2: Estimated 2025 operator latency-driven revenue, by application and latency level	18
Chart 3: Forecasted Mobile Operator low-latency revenue by application through 2025.....	19
Chart 4: Forecasted MNO low-latency revenue thru 2025, by world region	20
Chart 5: Forecasted data center deployment by proximity to users	23
Chart 6: Forecasted data center deployment by proximity to users.....	24
Chart 7: Forecasted data center deployment by proximity to users	25
Chart 8: Forecasted number of ‘Edge’ data centers in service	26
Chart 9: Forecasted edge server shipment, by data center proximity to end users	27
Chart 10: Forecasted edge server shipment, by indoor and outdoor	27
Chart 11: Forecasted edge server shipment, by world region	28
Chart 12: Forecasted edge server shipment, by data center business model	29
Chart 13: Forecasted server shipments, Virtual Core, VRAN, and EC	29

FIGURES

Figure 1 Edge Computing TCO vs. Revenue for various levels of latency.....	6
Figure 2 Edge Computing conceptual diagram	9
Figure 3 Edge Computing general trend.....	9
Figure 4 Latency for each step in a typical Mobile use of Cloud apps.....	10
Figure 5 Edge Computing applications, segmented by latency requirement.....	11
Figure 6 Latency vs Message Size for LTE and 5G	12
Figure 7 Conceptual diagram of REIT-hosted Cloud ecosystem	14
Figure 8 Assumptions for Edge Computing cost calculations.....	15
Figure 9 Edge Computing cost calculations for IoT case.....	16
Figure 10 Edge Computing cost calculations for Video CDN case	17
Figure 11 A typical setup for a national or regional data center.....	21
Figure 12 An example of a city data center in a steel container	22
Figure 13 Examples of outdoor climate-controlled enclosures.....	22



MOBILE EXPERTS

Edge Computing for Mobile Networks 2019

MEXP-EDGE-19
April 2019

Entire contents © 2019 Mobile Experts Inc.. Reproduction of this publication in any form without prior written permission is strictly forbidden and will be prosecuted to the fully extent of US and International laws. The transfer of this publication in either paper or electronic form to unlicensed third parties is strictly forbidden. The opinions expressed herein are subject to change without notice.

1 EXECUTIVE SUMMARY

Edge Computing is a major trend for the wired Internet, and investment has been tremendous for Content Delivery Networks and regional colocation centers for the past 10-20 years. For Facebook, Netflix, Snapchat, and YouTube, localization of content is a well-established fact. However, in the mobile telecom world, Edge Computing has barely begun.

The same driving forces exist in both markets. Edge Computing can reduce latency by locating the content or computing function closer to the end user. Also, video content that is cached at the edge can save millions of dollars on upgrades to long-distance transport capacity.

In the mobile market, these factors have only become important over the past 3 years, as most users did not consume appreciable amounts of video content locally, and IoT applications with low-latency requirements had not emerged. The rise of 5G makes Mobile EC important because the low cost/GB of 5G makes video delivery a mainstream application, and rapid IoT applications will emerge,

The definition of the “Edge” is very different for the Web-scale players, mobile operators, and enterprises:

- Web scale players deploy each regional data center to cover 200km or more, and refer to their regional data centers as “Edge Computing” sites.
- Mobile operators can deploy computing resources to their Central Office locations, typically within 100-200 km of the user. This is sometimes referred to as the “near edge”.
- Mobile operators also have the opportunity to deploy Edge Computing near the tower, or less than 20 km from end users. This can be called the “far edge”.
- Enterprises typically think about the “edge” as a deployment of computing resources on their premises, within 1 km of the client device, or even on the client platform itself.

Each of these definitions is valid for its respective market area, and each player will invest in different locations. This study is focused on investment related to the mobile network.

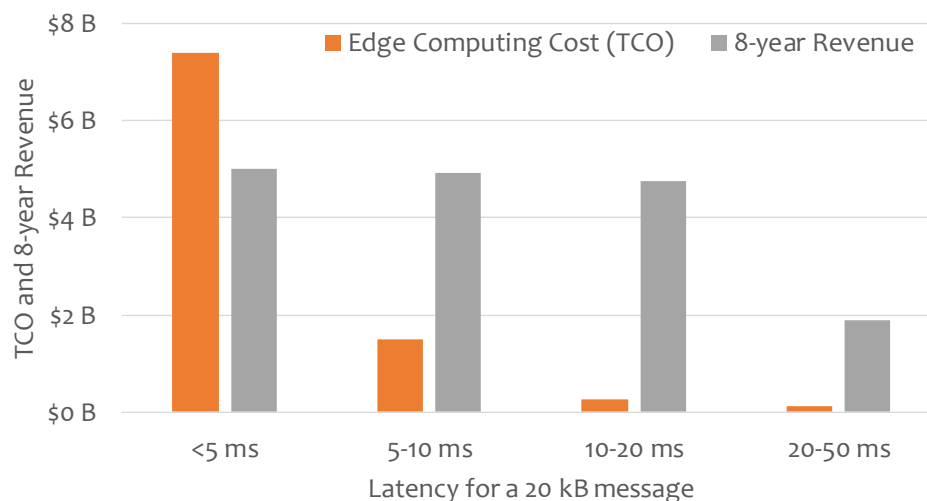


Figure 1 Edge Computing TCO vs. Revenue for various levels of latency

Source: Mobile Experts. TCO includes CAPEX and 8 years of OPEX for a nationwide US network. Revenue represents new latency-enabled revenue.

Deployment of a network-wide Edge Computing capability can be extremely expensive, considering the cost of renting space at a tower site and the maintenance of an advanced computing platform. Today, it seems clear that deployment of Mobile EC at every tower would not be justified by the added revenue. However, deployment of Mobile EC in regional data centers can be a good compromise in a market like the USA, where we see pent-up demand for low-latency wireless applications.

We see the best strategy for operators in the 10-20 ms target range, because the cost of deployment is fairly low while the operator can still address most of the currently visible revenue opportunities.

One critical insight here: The 20 ms target is best for mobile operators because the Web scale players will invest in regional data centers to support it... reducing the investment by MNOs. Chasing total latency below 10 ms will provide little payoff for the operators, at huge expense.

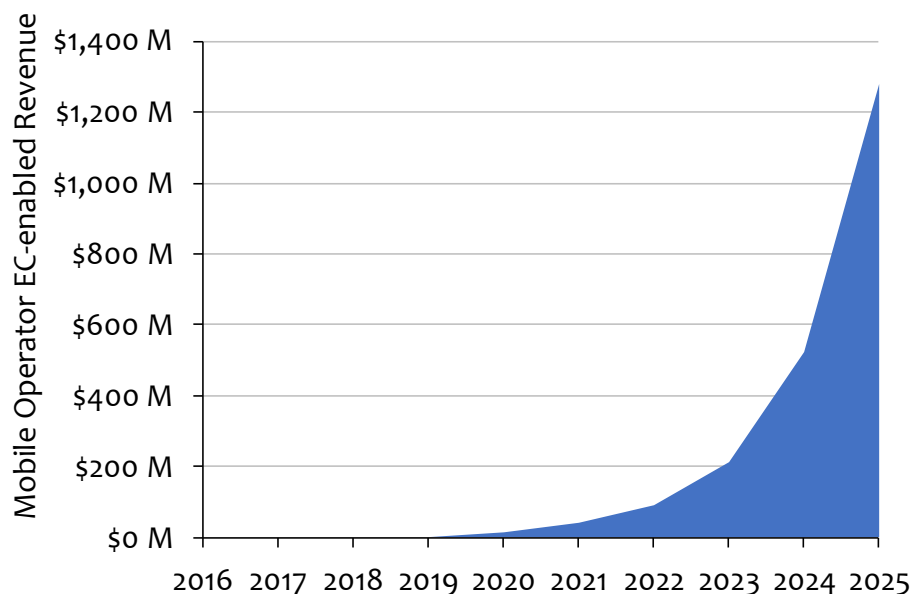


Chart 1: Forecasted Mobile Operator EC-enabled Revenue thru 2025

Source: Mobile Experts

Worldwide, we expect Edge Computing to settle into a business model that makes sense for multiple players. Mobile operators will uncover more than \$1B of new latency-sensitive revenue by 2025. Web-scale players will make billions of dollars on mobile advertising. Enterprises will save significantly through automation using low-latency robotics and other innovations.

In Mobile EC, we believe that simply showing a positive ROI for the operator is not adequate... all of the players in the chain must benefit. In some key applications, the enterprise will be strongly motivated to pay for premium wireless services, or to subsidize the Edge Computing deployment through subscription fees to a Cloud service. In this way, the Cloud players, colocation providers, mobile operators, and OEMs can all benefit simultaneously. Note that in other applications, the market will be too small for such a win-win-win scenario.

In particular we list the following conclusions for latency-sensitive applications:

- Consumer AR/VR will be a small market without enough critical mass to drive Mobile EC.
- Gaming is a big market that is likely to step directly into Mobile EC and drive significant revenue.
- Industrial robotics will be a small market that will be deployed locally, not nationally.
- HD Mapping for automotive applications will be useful for Level 3 automation in cars, and auto OEMs are likely to pay for Mobile EC support in some cities.
- Industrial Augmented Reality is getting some traction, and will use Mobile EC to improve the experience for the users.

2 MARKET DRIVERS AND CHALLENGES

In this study, we examine the direction for mobile operators to engage in edge computing. How will the business model change? What forces are pushing the mobile operators to adopt a different business model? What are the revenue opportunities and the big investments necessary?

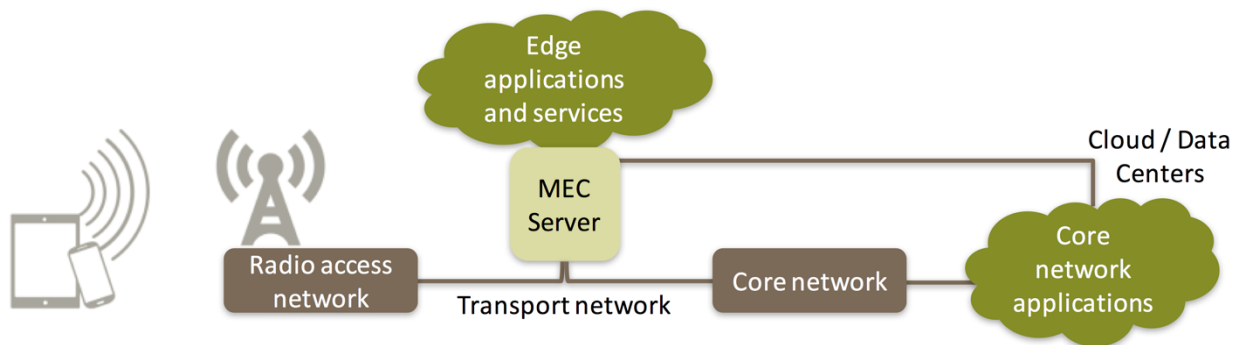
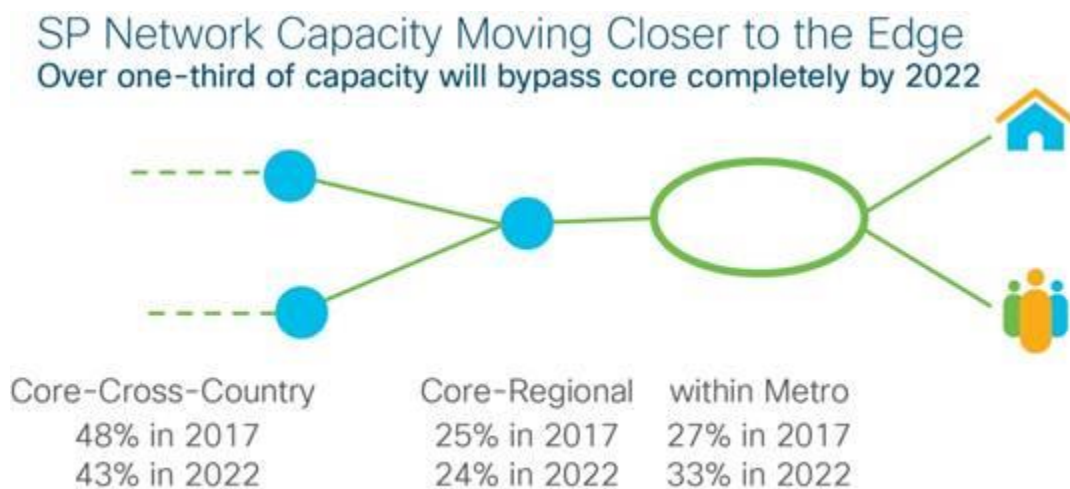


Figure 2 Edge Computing conceptual diagram

Source: Mobile_Experts

A great deal of work is taking place to move toward the Edge Computing architecture. In the wired Internet, we've already reached a point where more than half of data is served from regional data centers or cached locally. The wireless network is behind the adoption level of wired Internet in this regard, so clearly the mobile operators are looking to catch up.



Source: Cisco VNI Global IP Traffic Forecast, 2017-2022

Figure 3 Edge Computing general trend

Source: Cisco VNI

The Apps

A few key apps are driving the trend toward Edge Computing, and for a variety of different reasons. We outlined and analyzed each application in more detail in [last year's report](#), but for practical purposes we can focus on two primary benefits: Lower Latency and Distributed Capacity.

Lower Latency

Lower Latency comes from a combination of reduced radio latency and localized computing. In fact, the localized compute resource makes a much bigger difference than a new radio interface when it comes to improvements in latency. So, the use of 5G to reach 1 ms latency in the radio will do nothing on its own...it must be paired with on-site Edge Computing to reach the sub-10 ms overall latency required for some applications.

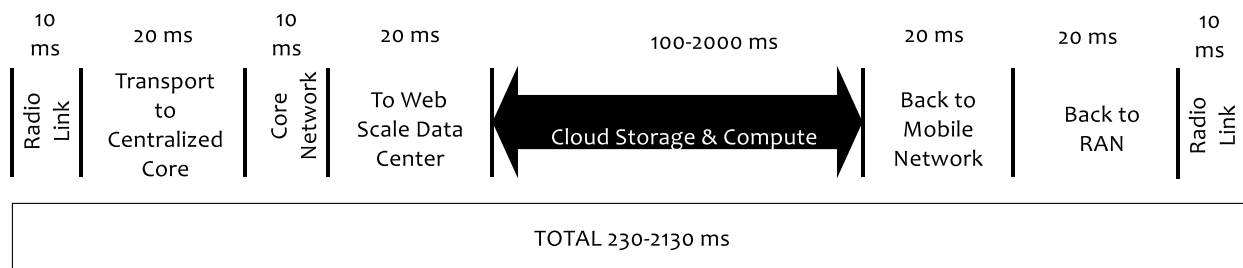


Figure 4 Latency for each step in a typical Mobile use of Cloud apps

Source: Mobile Experts

Distributed Capacity

The other major driving force for Edge Computing is the cost of long-distance transport. In particular, with streaming video the content providers are moving to place the content on the edge of the network, with regional CDNs so that latency is reduced for their app. In general we believe that this will drive improvement for the operators without the MNO needing to invest in a CDN of their own at the radio tower. So, the move toward Edge Computing is happening but it's driven by Netflix and Youtube, and less by Verizon or NTT DoCoMo.

For convenience, here is a quick summary of how we view the apps by latency requirement:

Total Latency	Applications
≤ 5 msec	<ul style="list-style-type: none"> • Telepresence with real-time synchronous haptic feedback • Industrial robots • Closed-loop industrial control systems • Negotiated automatic cooperative driving • Smart grid: Synchronous phasing of power supplies
≤ 10 msec	<ul style="list-style-type: none"> • Shared haptic virtual environment • Tele-medical applications, e.g. tele-diagnosis, tele-rehabilitation • Augmented reality • Haptic overlay trainer/learner for fine motor skills (e.g. for medical) • Smart grids • Process automation
≤ 50 msec	<ul style="list-style-type: none"> • Serious gaming (20 msec) • Cognitive assistance (20-40 msec) • Virtual reality • Cooperative driving (20 msec) • UAV control (10 – 50 msec) • Remote robot control with haptic feedback (25 msec) • Automotive pre-crash sensing warning
≤ 100 msec	<ul style="list-style-type: none"> • Vehicle safety apps (mutual awareness of vehicles for warning/alerting) • Assisted driving – cars make cooperative decisions, but driver stays in control

Figure 5 Edge Computing applications, segmented by latency requirement

Source: Mobile Experts

Here's the interesting part about latency: The radio latency matters for very small messages, but for larger messages the latency is much more dependent on the data speed, not the "ping" latency. As an illustration, consider an automotive application where HD maps are loaded into the car's memory. When the car approaches a hazard, it detects the hazard and uploads images so that the network can quickly pass the new HD map data to cars following behind. If the file size is more than 1 MB, then the location of edge computing doesn't matter very much, but the data speed of the connection is critical. In this safety-related scenario, the mapping application will work best if it can quickly deliver a very light message with the location of the hazard, followed by more detailed data.

In the case of industrial robotics, it's clear that with either wires or wireless the control messages must be very light (below 20 kB) in order to achieve the sub-5 millisecond latencies required.

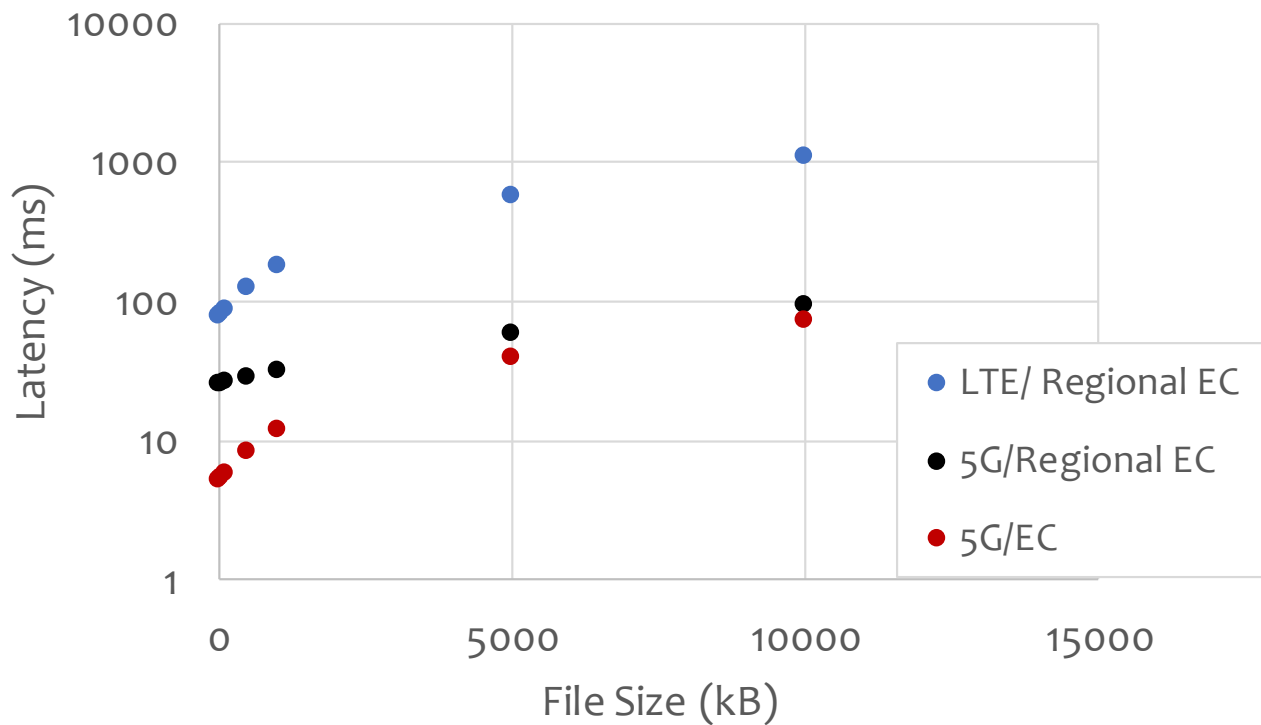


Figure 6 Latency vs Message Size for LTE and 5G

Source: Mobile Experts

The Edge Computing business model

As the wide market considers multiple applications, there are multiple business models that could come into play, to provide the low-latency local computing necessary for future growth:

- The mobile operators could invest in creating their own Cloud;
- The operators could host the web-scale Cloud players in their Central Office or other locations in each city;
- REITs could step in to prepare edge computing sites in the best locations in each market;
- The web-scale Cloud players could establish hundreds to thousands of their own data centers to cover each city.

Let's look at each of these options for signs of life:

... Mobile operators and a "Telco Cloud"

This option was seriously considered by leading operators in the 2010 through 2018 timeframe. The world's top 20 operators dabbled in setting up their own data centers and cloud infrastructure, in hopes that enterprises or consumer app developers would invest in services that operate on the Telco Cloud.

This fantasy never materialized. Verizon sold 29 data centers to Equinix in 2017, and AT&T recently closed a deal to sell 31 data centers to Brookfield/Evoque. Vodafone, Tata, and other global operators have made similar moves at a smaller scale.

The fundamental problem here is that enterprises and app developers don't want to develop for the "Verizon Cloud" or the "Vodafone Cloud". In particular, large multinational companies prefer to work with a global web-scale player such as AWS or Microsoft Azure, for continuity and predictable performance across multiple continents.

... Operator-hosted Cloud

The mobile operators have two key assets: They own wireless spectrum and they control real estate in every major city. In theory, the operators could host the web-scale Cloud players to offer low-latency services.

This one is still a possibility, but we have serious doubts. Mobile operators don't like to share anything, and as an example, if one operator hosts AWS in a local data center, then a second or third mobile operator could be excluded from gaining access to the data center. That doesn't work for the web-scale guys. We believe that AWS, Google, and Microsoft would rather not support three separate telco-hosted cloud centers in each city.

... REIT hosted data centers

This business model is getting some serious investment now. Companies such as Cyxtera, Vapor IO, EdgeConnex, and others are rapidly building out data centers to host multiple Cloud players in each location. AWS and Microsoft Azure already like to work in this type of colocation business model, so the top two web-scale players are already on board. Google acts more independently, but the flexibility of a plain-vanilla colocation facility is that anyone (including the operators themselves, or OTT players such as Netflix/Facebook/Snapchat) can enter easily.

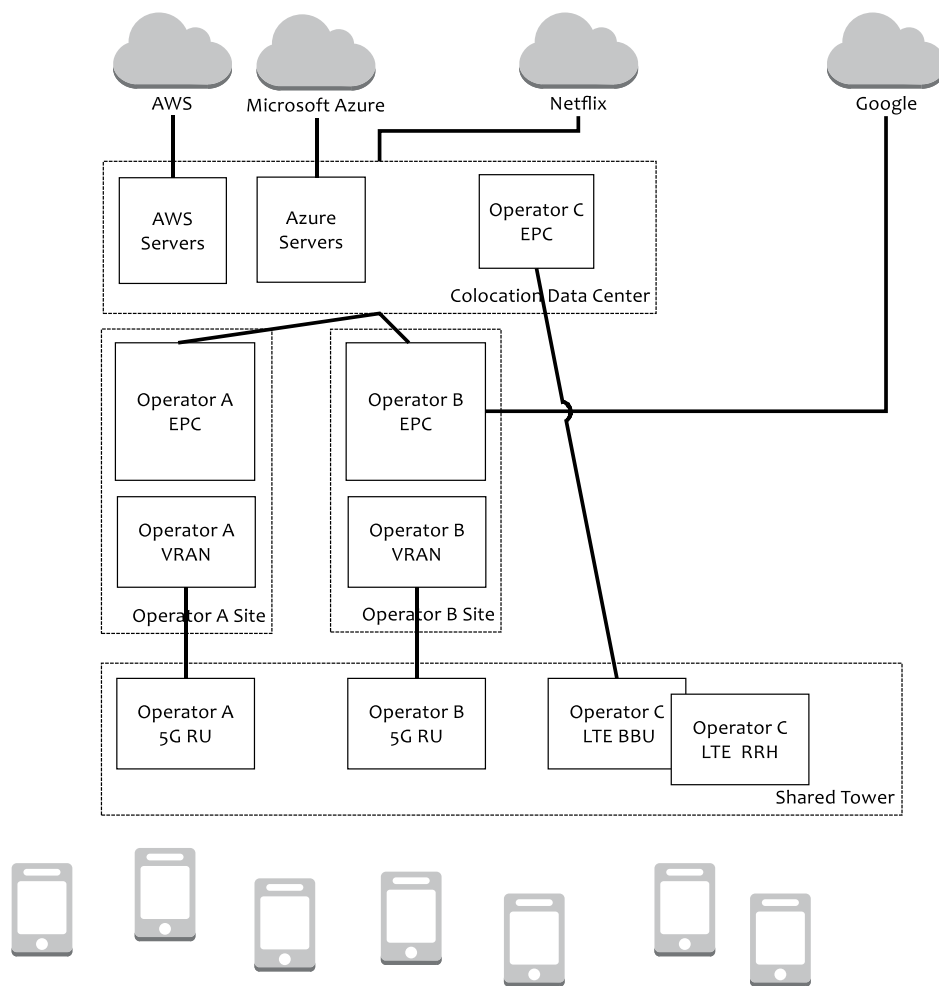


Figure 7 Conceptual diagram of REIT-hosted Cloud ecosystem

Source: Mobile Experts

... Web-scale players deploying their own data centers

Rounding out the field, of course we must consider the option that all of the Web scale players will invest in their own local data centers to differentiate. This is the approach taken by Google, Facebook, and some other Cloud players that prefer to control their own infrastructure and security. In general the trend is toward co-location because for many players the duplication of investment would be enormous, so we see strong economic leverage to share costs by colocating.

The economics of Edge Computing: does it save money?

The economics of Edge Computing can be confusing, as every Edge Computing application puts a different kind of load on the network and as a result, the pricing models are complex.

Let's look at an example: For a hypothetical campus with 100,000 IoT sensors for security, lighting, and fire alarms, a university wants to control everything through the cloud. For simplicity, we assume that each sensor sends one 50 byte reading every second.

Assumptions	Centralized Cloud	Edge Cloud
IoT Sensors	100,000	100,000
Readings/second	1	1
Bytes / reading	50	50
GB / month	216	216
% handled locally	0%	90%
Shadow Requests per device per day	100	100
Rules/Actions triggered per device per day	50	50

Figure 8 Assumptions for Edge Computing cost calculations

Source: Mobile Experts

Over an 8-year period, the costs add up despite the low cost to support each individual sensor. In the online pricing offered by AWS, we expect five main contributors to Cloud Computing costs:

1. Transport costs for long-distance connection to a centralized AWS data center;
2. Connectivity costs to have a connection available for our 100,000 devices;
3. Messaging costs at \$1.20 per billion messages;
4. Device shadowing and registry charges of \$1.25 per million ‘events’; and
5. Charges for each rule or action triggered to send a message.

Costs	Centralized Cloud	Edge Cloud
AWS Transport cost (\$/GB)	\$0.04	\$0.04
Monthly Transport Cost	\$8.64	\$0.86
AWS cost per million minutes of connection	\$0.08	\$0.08
AWS Cloud cost per month	\$345.60	\$34.56
AWS cost per billion messages	\$1.20	\$1.20
AWS Cloud cost per month	\$311.04	\$31.10
AWS cost per million shadow/registry events	\$1.25	\$1.25
AWS shadow/registry cost per month	\$ 375.00	\$ 37.50
Cost per million rules or actions triggered	\$0.15	\$0.15
AWS rules engine charge/month	\$22.50	\$2.25
Total Cost / month	\$1,062.78	\$106.28
Cost of Server	\$0	\$25,000
Maintenance cost/month	0	\$200
Total 8 year Cost of Ownership	\$102,027	\$54,403

Figure 9 Edge Computing cost calculations for IoT case

Source: Mobile Experts

In the end, for this case it's cheaper to buy a server and operate it on-site. A large university may choose to make the investment if they have an IT staff that's capable of running its own server... but many smaller businesses will not see enough savings here to justify taking on this challenge. A smaller community college is much more likely to stick with centralized cloud computing.

Five years ago, the drive toward Mobile Edge Computing seemed sensible based on the 2014 economics of long-haul transport and expectations for low-latency IoT applications. But the web-scale market did not stay constant, and the low-latency wireless market failed to appear as quickly as the pundits expected.

In 2014, some analysts reported the cost of long-distance transport for IP data in the range of \$1-2 per GB. That number was used to justify investment toward Edge Computing, especially focused on video caching. The idea was that caching video at the edge would save millions of dollars in transport costs. However, today the operators don't see such a compelling ROI. Costs have

dropped to the range of \$0.04/GB (regardless of distance within the USA), and only \$0.13 per GB from the USA to Australia.

Assumptions	Centralized Cloud	Edge Cloud
Video customers	1,000,000	1,000,000
GB/month/user	180	180
% of common videos	30%	30%
GB of video cached (avg per customer)	0	54
GB of video for long-distance transport	180,000,000	126,000,000
GB of video stored in cache (total)	540	540
Costs	Centralized Cloud	Edge Cloud
AWS Transport cost (\$/GB)	\$0.04	\$0.04
Monthly Transport Cost	\$7,200,000	\$5,040,000
CDN Costs (\$/GB/month)	\$0.20	\$0.50
Monthly Cloud Cache Cost	\$108.00	\$270.00
Total Cost / month	\$7,200,108	\$5,040,270
Total 8 year Cost of Ownership	\$691,210,368	\$483,865,920

Figure 10 Edge Computing cost calculations for Video CDN case

Source: Mobile Experts

So, there is an incremental savings that is possible if the video streaming service can set up a local CDN and reduce the traffic transported between a centralized cloud location and their local area. This is not a dramatic savings, so we hear some disagreement among different vendors with regard to the importance of localizing video. One concrete example is that Verizon agreed to move 20% of its video traffic to CDNs (with Qwilt as a vendor) during 2017.

The economics of Edge Computing: does it add revenue?

The other side of the ROI equation is revenue... and in the case of Edge Computing, the primary question is whether lower latency can create new revenue. Multiple applications have been proposed:

- Robotics/Manufacturing
- Heavy Equipment Automation (Mining/Seaports/Trains)
- Automotive Safety
- Drone Controls
- Virtual Reality
- Gaming

In the past, mobile operator revenue has always been based on speed and data capacity. People pay for a phone line and 10-20 GB of data. We expect a new profile to emerge over the next five years, where operators will derive revenue from enterprise customers, based on the latency performance of the network.

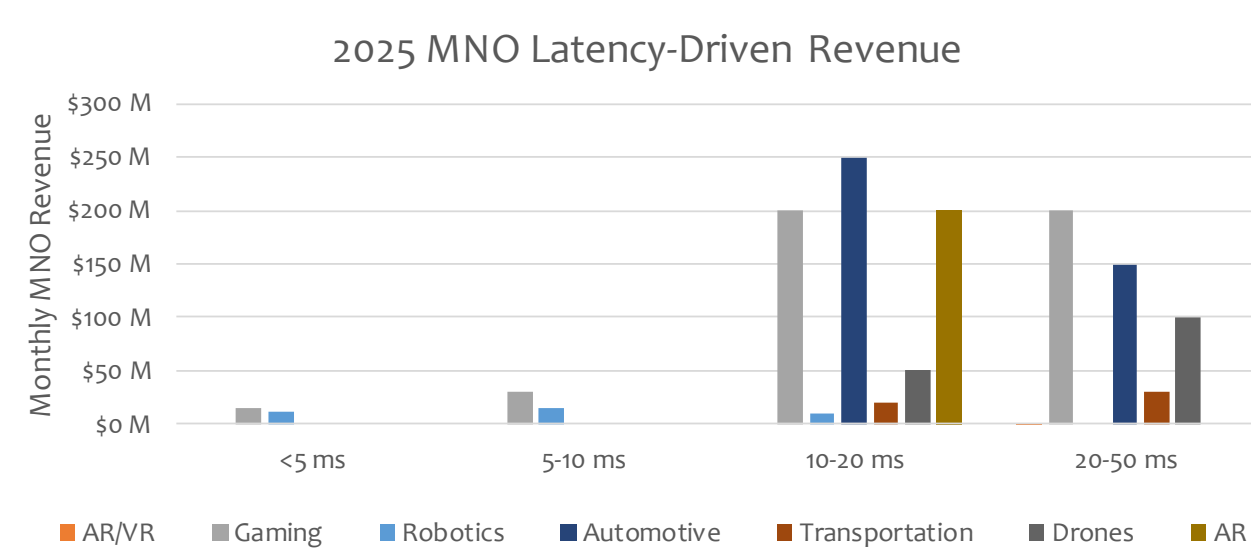


Chart 2: Estimated 2025 operator latency-driven revenue, by application and latency level

Source: Mobile Experts

Based on our interviews, we expect the 10-20 ms segment to generate the most significant revenue, led by gaming in the consumer segment as well as automotive, other transportation, and AR applications. Factory robotics does not factor into the revenue picture for operators as highly as expected, because factory automation is expected to be implemented as an on-premises Private LTE/5G network...thus resulting in no revenue for the operator aside from possible spectrum sub-leasing.

Total latency-driven revenue in 2025 is estimated to be roughly \$20 billion for mobile operators worldwide. To address the opportunity, mobile operators will need to invest significantly in edge computing infrastructure over the next five years.

3 REVENUE OUTLOOK

The revenue outlook for mobile operators is critical to the success of Edge Computing within the mobile market. The bottom line: if operators don't reach enough revenue, the EC trend will continue with web-scale players at a regional level but will not offer any new leverage for the mobile operators.

Mobile Operator Revenue

The operators are currently exploring dozens of applications, and in our interviews the revenue picture is extremely unclear to them. Almost zero revenue is available today for MNOs related to low-latency services...but all major operators are considering data plans and services based on low-latency guarantees with Edge Computing.

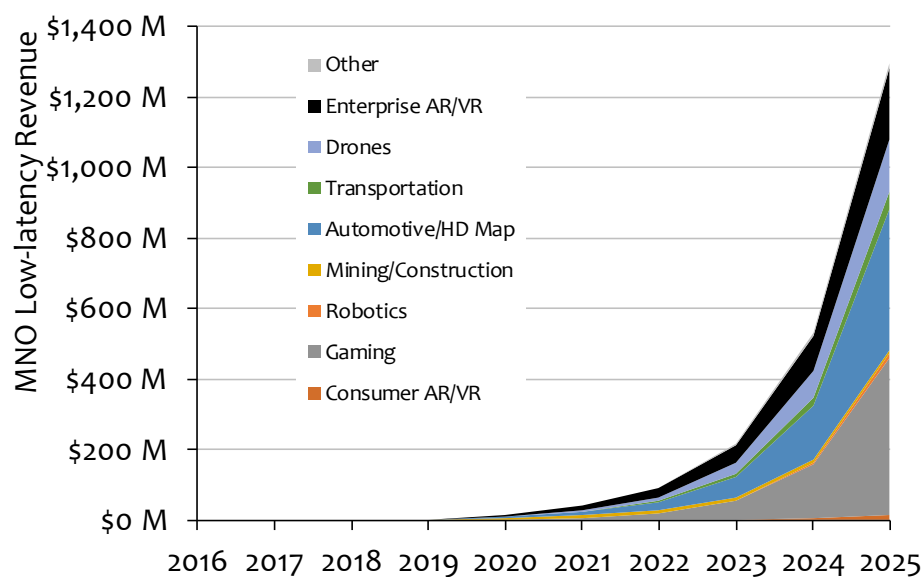


Chart 3: Forecasted Mobile Operator low-latency revenue by application through 2025

Source: Mobile Experts

The leading application will be consumer gaming, where hard-core gamers are likely to pay up to \$30 per month for a low-latency connection. (The gaming application may also succeed even without any guaranteed level of latency). The automotive market may also be a hit, as HD map updates must happen within about 50 ms in some cases, and below 20 ms in other emergency cases. Industrial Augmented Reality is a third leading prospect, as both throughput and latency are important to the usefulness of the glasses for modeling, repair, and complex industrial manipulation.

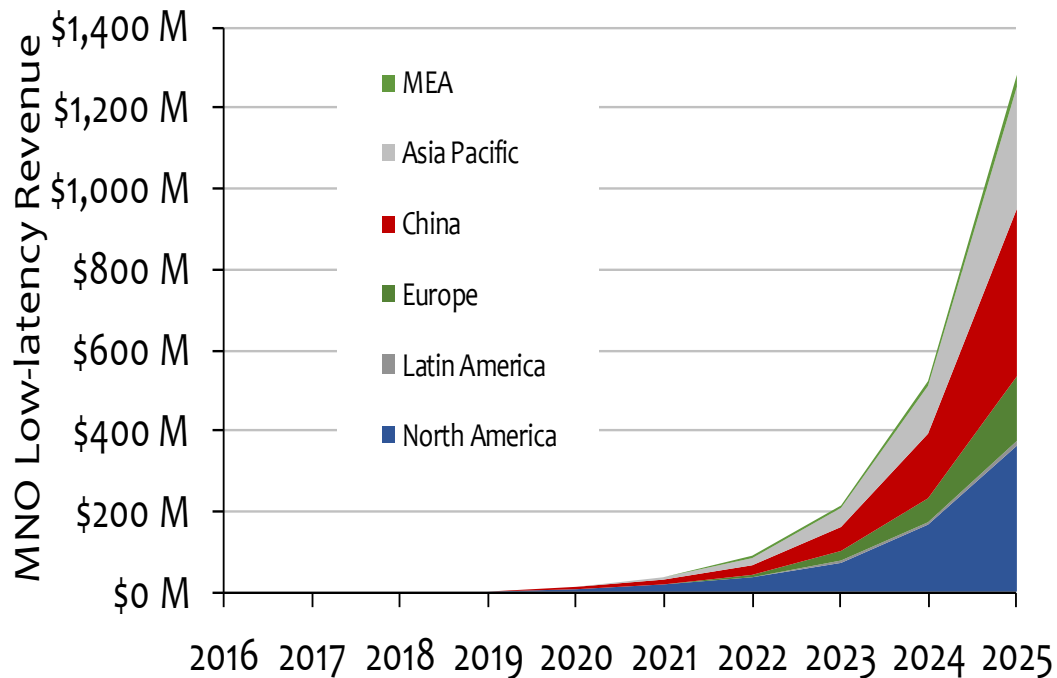


Chart 4: Forecasted MNO low-latency revenue thru 2025, by world region

Source: Mobile Experts

North America is the key market for early development of Edge Computing in both consumer and enterprise applications...but we expect China to also represent a significant market. In China, the government-owned operators are expected to invest directly in EC resources and we expect industrial AR and robotics applications to be prominent. Korea, Japan, and northern European countries are expected to also participate in the MNO-based EC market. (Note that each of these regions also has its own level of adoption for Private LTE/5G and related EC, which will compete with MNO-based EC. North America, Germany, and Finland in particular will see significant investment in on-premises EC using Private LTE/5G on private spectrum.

4 DATA CENTER AND SERVER OUTLOOK

This high-level forecast is intended to identify the likely investments in data centers, including their probable location and features. The following charts focus on the global deployment of data centers to serve mobile networks (LTE and 5G). The intent is to focus on support of mobile network operators... but of course many of these data centers serve wired networks, enterprise networks, and many other purposes. In the end, the Mobile Experts forecast for “Edge Data Centers” includes data centers deployed by MNOs directly, and data centers set up by hosting companies that also serve many other markets. Only the dedicated data centers for web-scale players or non-mobile telecom players are excluded.

Definitions

In this emerging field, the terminology is not well understood between multiple players. Where is the “edge”? What constitutes a “data center”?

For the purposes of our forecast, we have set up the following definitions:

Data center: Any location where standard off-the-shelf servers can be deployed to perform computing and storage functions.

Wide Area/Colocation data center: Typically a large building, 20,000 square feet or more, located hundreds of kilometers away from the end users supported. Measured by capacity, these sites can be over 100 MW.

Regional data center: A second-level location that extends the data center function to within 200 km of end users. These locations typically consist of buildings up to about 20,000 square feet and capacity ranges widely, from 50 MW to 150 MW or so.



Figure 11 A typical setup for a national or regional data center

Source: Terragraph

City data center: A third-level location, located within 20 km of supported end users. These locations are often implemented with container-style enclosures instead of fixed buildings. Typical capacity is less than 300 kW.



Figure 12 *An example of a city data center in a steel container*

Source: Vapor IO

On-site data center: Can be a rack of equipment in an enterprise or a ruggedized outdoor enclosure at the base of a radio tower. In this case the end users are typically within 2-3 km, and capacity would be less than 50 kW in most cases.



Figure 13 *Examples of outdoor climate-controlled enclosures*

Source: Rackmount Solutions

Data Center Forecast

To begin, we estimate that the number of data centers deployed at “the edge” in 2019 will number about 500 worldwide, counting new regional data centers in second-tier markets (most of which are colocation facilities for web-scale servers), as well as smaller city-level data centers such as the containers deployed by Vapor IO.

Growth of the edge data center market will continue steadily for at least 10 years, as web scale players distribute their networks more widely and push to reduce the lag time for customers. Based on the deployment plans for containerized data centers (5-10 locations per major city), we estimate that numbers could reach a few thousand per year.

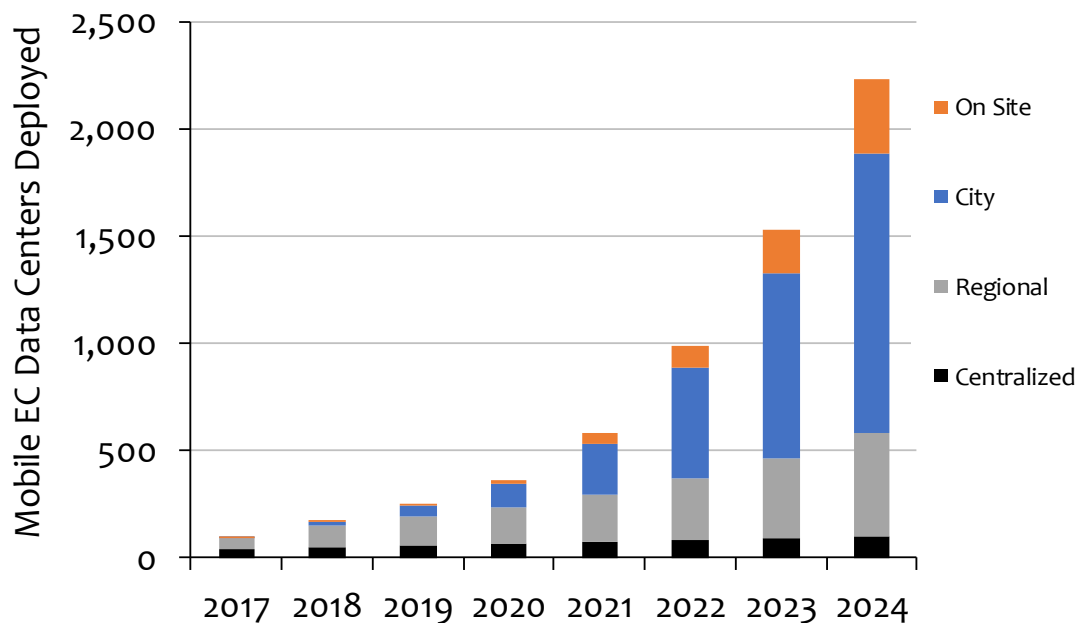


Chart 5: Forecasted data center deployment by proximity to users

Source: Mobile Experts

Most data centers today are located in buildings with giant air conditioners...and that approach will continue in the case of national-level and regional data centers. The smaller sites closer to the eds are moving toward containerized shelters...and will eventually include air-conditioned outdoor enclosures that are similar to the old telecom enclosures.

On-premises deployments may be rack-mounted equipment... for example, when an operator implements an in-building network for a factory, we expect them to have an indoor location with air conditioning that is suitable for a few racks of network equipment as well as edge computing server racks.

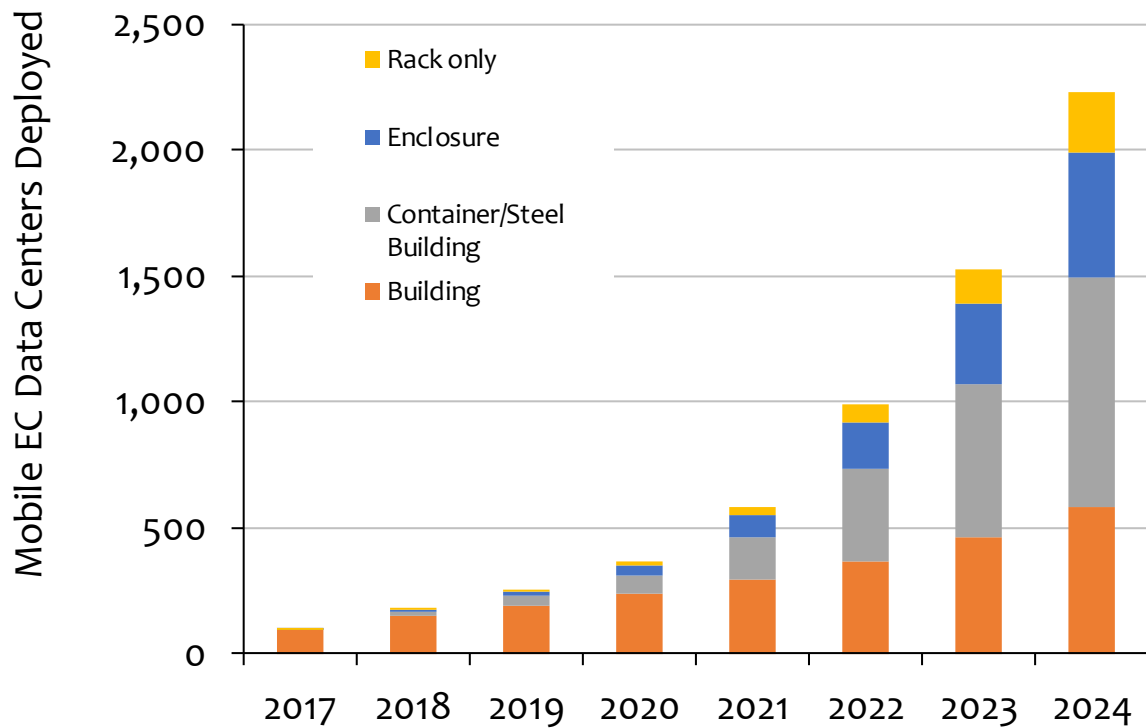


Chart 6: Forecasted data center deployment by proximity to users

Source: Mobile Experts

Mobile Operators will get involved in a wide variety of computing configurations, including use of web-scale data centers in regional colocation facilities, as well as more local facilities either hosted by neutral companies or in their own Central Office locations. Some enclosures will be right at the tower itself, or on the enterprise premises.

We anticipate that most applications will run on Cloud servers run by AWS, Microsoft Azure, and Google...and these companies primarily rely on regional colocation centers. Over time, these web-scale companies should start to collocate into city-level data centers as well, taking advantage of “micro” locations offered by companies such as Vapor IO, Baselayer, and EdgeMicro.

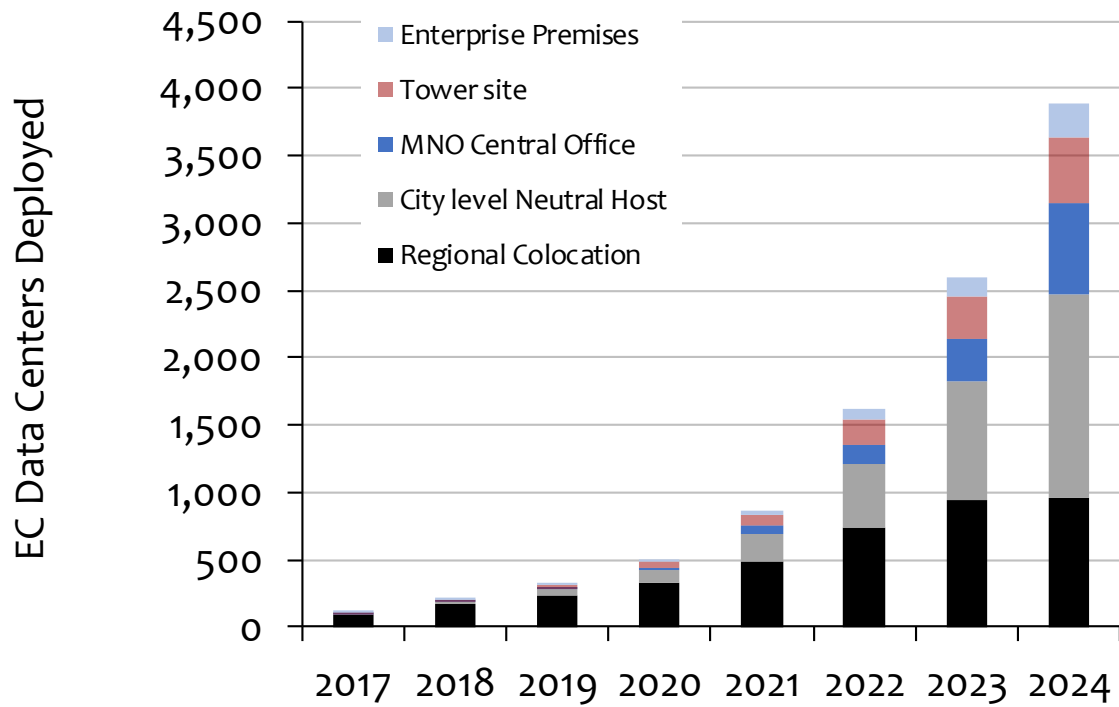


Chart 7: Forecasted data center deployment by proximity to users

Source: Mobile Experts

The ‘installed base’ of edge data centers will grow steadily into the thousands, as in the near terms the market will move toward roughly one data center per million users...and then toward one edge data center per half-million users.

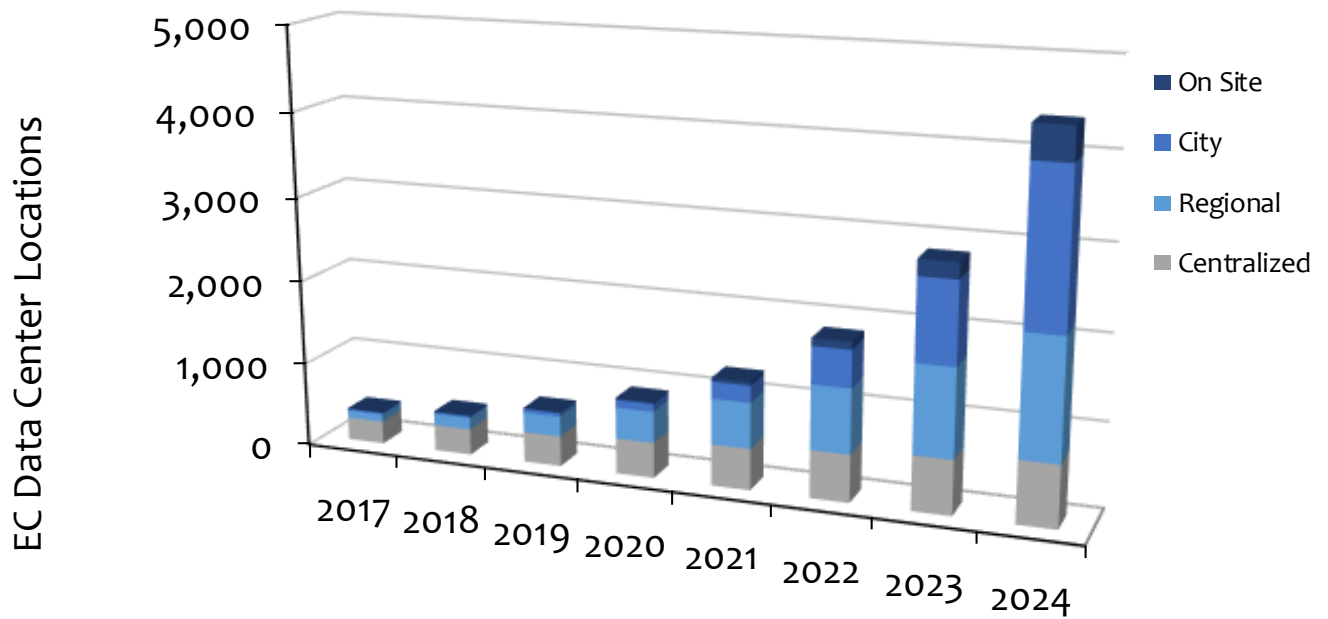


Chart 8: Forecasted number of 'Edge' data centers in service

Source: Mobile Experts

Edge Server Forecast

The massive data farms run by companies like Google (millions of square feet and over 100,000 servers each) are now migrating toward smaller, more distributed data centers in colocation facilities. Each web-scale player has a different strategy here, but the general trend is toward smaller numbers of servers per site.

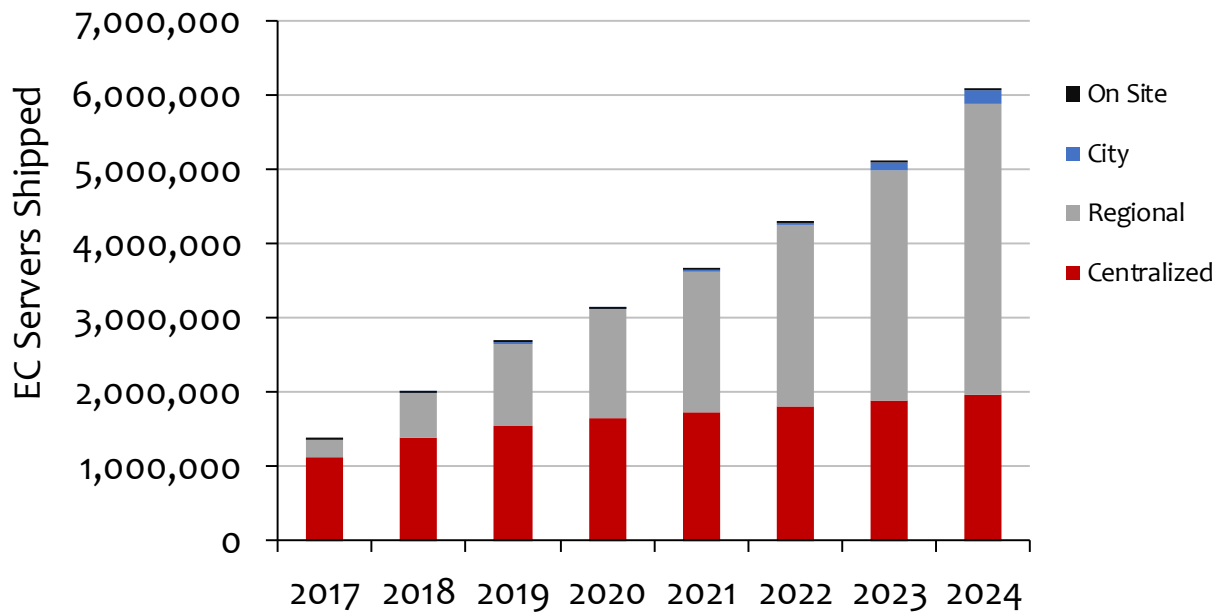


Chart 9: Forecasted edge server shipment, by data center proximity to end users

Source: Mobile Experts

We're tracking the form factor of the servers themselves, to watch for deployment of standard X86 silicon in a ruggedized outdoor form factor. Will an outdoor version of the white-box server become a normal item? So far we don't see it, but we're tracking this possibility.

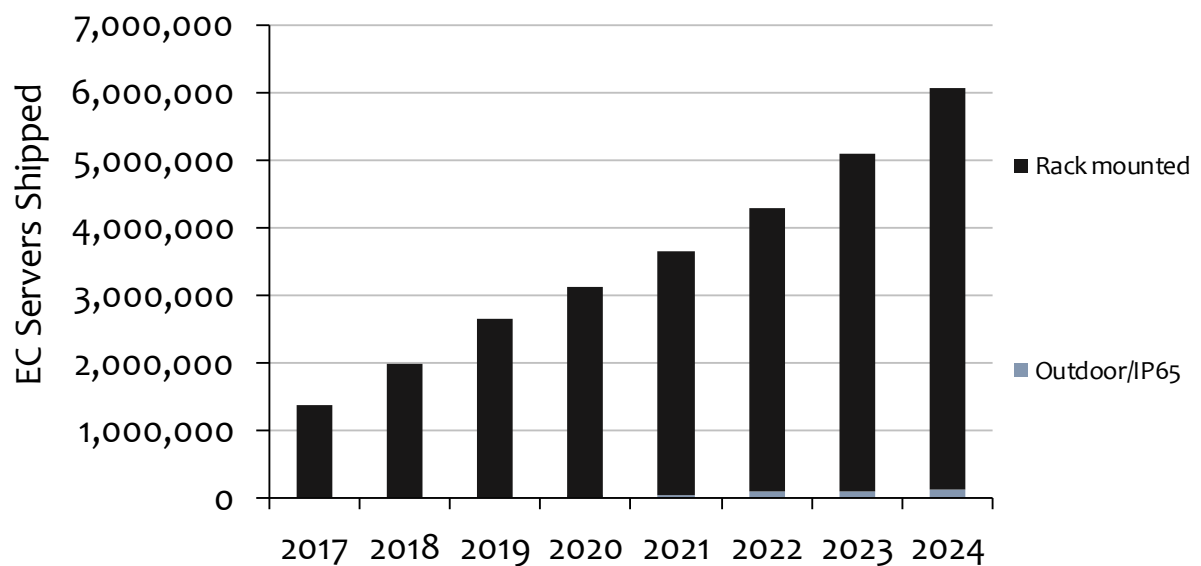


Chart 10: Forecasted edge server shipment, by indoor and outdoor

Data center capacity is growing quickly in China, USA, Japan, and a few other key markets. Note that it's more accurate to track the regional deployment of edge computing by server count, not by data center numbers... so we use this view to illustrate the regional breakdown.

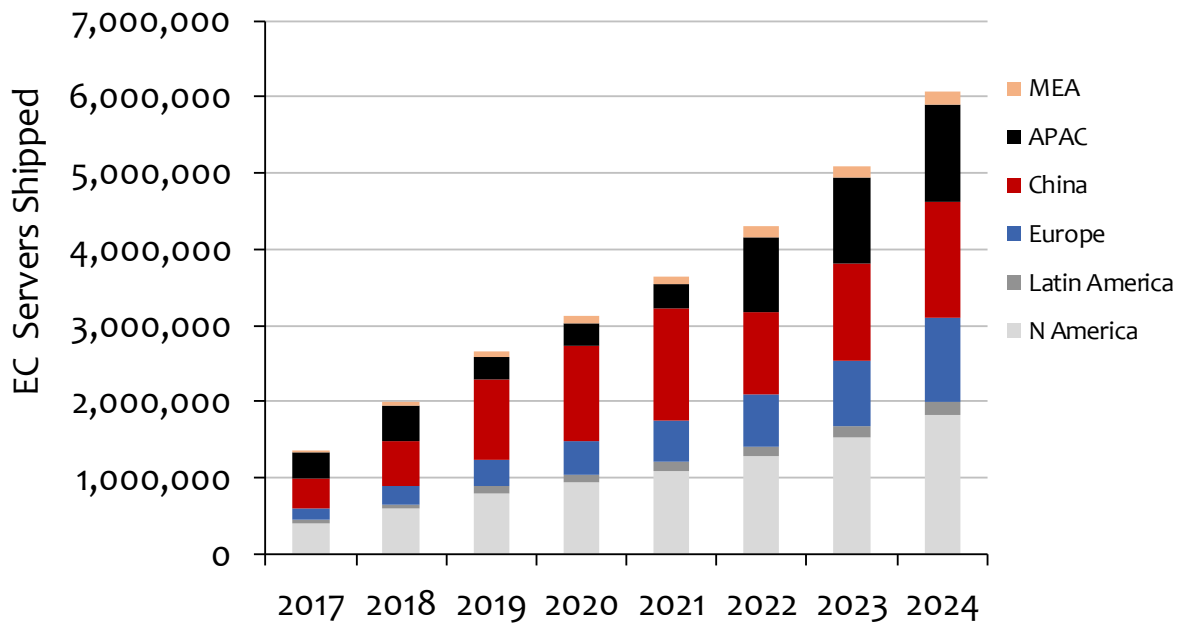


Chart 11: Forecasted edge server shipment, by world region

Source: Mobile Experts

The business model is migrating from servers in owned data farms to a collocated market where regional companies step up to host the web-scale players for localized services. Note that—as much as the telcos would like to run their own cloud in the convenience of their Central Office—the web scale Cloud systems will control the software in this market, and therefore they will dictate the business model.

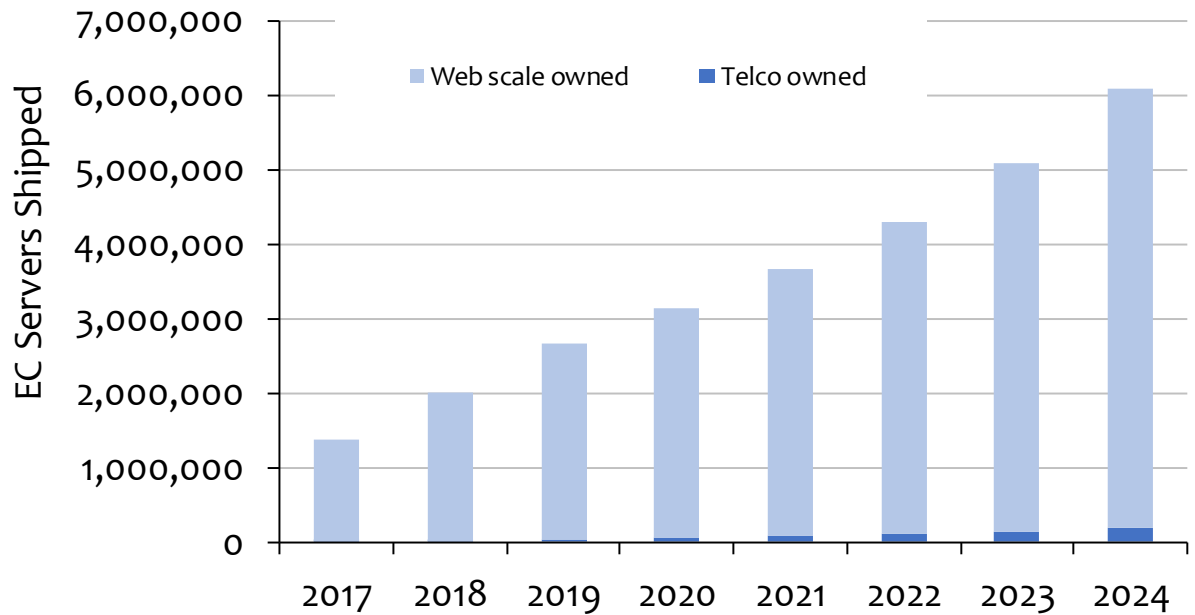


Chart 12: Forecasted edge server shipment, by data center business model

Source: Mobile Experts

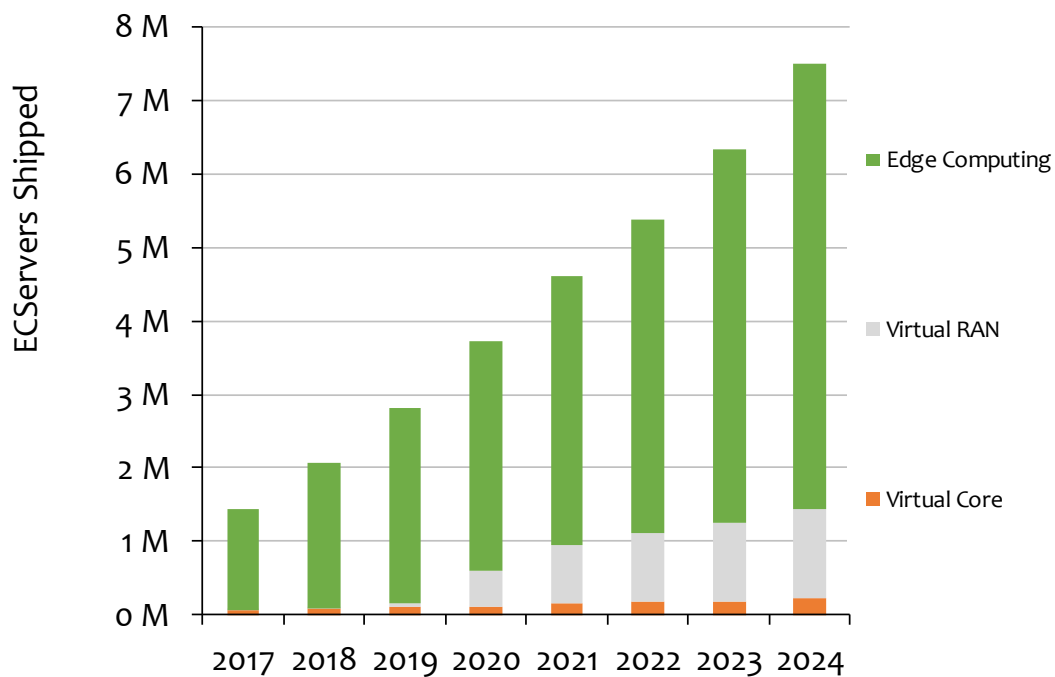


Chart 13: Forecasted server shipments, Virtual Core, VRAN, and EC

Source: Mobile Experts

5 KEY COMPANIES

CLOUD PROVIDERS

Alibaba

AWS (Amazon Web Services)

Google (GCP, Google Cloud Platform)

IBM

Limelight Networks

Microsoft Azure

Packet

DATA CENTER/COLOCATION

American Tower

CBRE

Colt

Compass

Crown Castle

Cyrus One

Digital Fortress

Digital Realty

EdgeConneX

Equinix

Global Switch

Hitachi

iseek

KVH

KDDI

Keppel

NextDC

Rackspace

SRP

Tata

ENCLOSURES/MICRO DATA CENTERS

Baselayer

DartPoints

DataBank

EdgeMicro

Modular Life Solutions/EdgePod

Stulz

Vapor IO

Vertical Bridge

CDNs

Akamai

CDN77

Cloudflare

Fastly

Google Cloud CDN

Highwinds

Imperva Encapsula

Qwilt

StackPath

HARDWARE OEMS

ADLINK

Cisco

Dell

Ericsson

HPE

Huawei

Intel

Nokia

NVIDIA

Samsung

ZTE

SOFTWARE SUPPLIERS

Affirmed Networks

Altiostar

Aricent

Clearblade

CPLANE.ai

Druid

Juniper Networks

Mavenir

Quortus

Rigado

RTI (Real-Time Innovations)

Saguna

Vasona Networks

VMware

6 METHODOLOGY

For our analysis of Edge Computing, we first focused the scope of the report on investment in Edge Computing data centers. In the Cloud Computing market, huge investment is taking place that is not necessarily tied to mobile networks, and in fact a great deal of the usage takes place on wired broadband networks. We focused our assessment of data center investments on the collocated/edge deployments of data centers where local breakout of a mobile network will be enabled and “Edge Computing” will be superior to simpler cloud computing at the core network site.

We interviewed 20 different Edge Computing players, ranging from web-scale companies to colocation specialists, micro-data center suppliers, and mobile operators. By checking in with multiple players in the ecosystem we hope to ensure an accurate view of the key drivers in the market.

The forecasts in this report are global in scope, but limited to data centers and servers that support mobile customers through the mobile network. This does not mean that all of the facilities we forecast will be exclusively used for mobile customers. Indeed, we expect that the “edge” colocation centers will be shared between wireless and wired networks as the Cloud moves to multiple edges at the same time.